Amendments to the Claims:

Listing of Claims:

1. (Currently amended) A method for constructing a variant set for modifying a biopolymer of interest, the method comprising:

a) identifying a plurality of positions in said biopolymer of interest and, for each respective position in said plurality of positions, one or more substitutions for the respective position, wherein the plurality of positions and the one or more substitutions for each respective position in the plurality of positions collectively define a biopolymer sequence space;

b) selecting a first plurality of variants of the biopolymer of interest thereby forming a variant set, wherein said variant set comprises a subset of said biopolymer sequence space;

c) measuring a property of all or a portion of the variants in the variant set; and

d) modeling, using a suitably programmed computer, a sequence-activity relationship between (i) one or more substitutions at one or more positions of the biopolymer of interest represented by the variant set and (ii) the property measured for all or the portion of the variants in the variant set, wherein the sequence-activity relationship has the form

$$Y = f(w_1x_1, w_2x_2, \ldots w_ix_i)$$

wherein,

Y is a quantitative measure of the property;

x_i is a descriptor of a substitution, a combination of substitutions, or a principal component of one or more substitutions, at one or more positions in the plurality of positions;

w_i is a weight applied to the descriptor x_i; and

f( ) is a mathematical function,

and wherein the modeling comprises:

i) optimizing, using a suitably programmed computer, the sequence-activity relationship by adjusting individual weights w_i for each said descriptor x_i using a refinement algorithm that minimizes the difference between the predicted values and the real values of Y from partial data, wherein the partial data is the first plurality of variants with either (1)

individual sequences left out on a random basis or (2) individual substitutions at positions in the plurality of positions left out on a random basis, and

ii) repeating the optimizing i) a plurality of times thereby obtaining, for each respective substitution or combination of substitutions x_i, (a) an average value for the weight w_i describing a relative or absolute contribution of the respective substitution or combination of substitutions x_i to Y, and (b) a standard deviation, variance or other measure of confidence in the weight w_i describing the relative or absolute contribution of the respective substitution or combination of substitutions x_i to Y.

2-116 (Cancelled)

117. (Previously presented) The method of claim 1, the method further comprising:

e) defining a new variant set for the biopolymer of interest that comprises variants that include substitutions in the plurality of positions that are selected based on a function of the sequence-activity relationship.

118. (Previously presented) The method of claim 117, the method further comprising:

f) measuring a property of all or a portion of the variants in the new variant set.

119. (Previously presented) The method of claim 1, wherein the plurality of positions and the one or more substitutions for each respective position in the plurality of positions are identified using a plurality of rules.

120. (Previously presented) The method of claim 119, wherein the plurality of rules comprises two or more rules selected from the group consisting of:

(i) the favorability of a substitution calculated from a substitution matrix;

(ii) the probability of a substitution calculated from a conservation index;

(iii) the proximity of a position to a structurally defined region within the biopolymer,

(iv) the presence of a substitution in a homologous biopolymer;

(v) the favorability of a substitution calculated from a comparison of homologous sequences;

(vi) the mutability of a position calculated from a comparison of homologous sequences;

(vii) the favorability of a substitution calculated from a comparison of homologous structures; and

(viii) the mutability of a position calculated from a comparison of homologous structures.

121. (Previously presented) The method of claim 1, wherein the variant set is enriched for pairwise uniqueness of substitutions at positions in the plurality of positions.

122. (Previously presented) The method of claim 1, wherein the variant set consists of fewer than 1000 variants.

123. (Previously presented) The method of claim 1, wherein the variant set consists of fewer than 250 variants.

124. (Previously presented) The method of claim 1, wherein the variant set consists of fewer than 100 variants.

125. (Previously presented) The method of claim 1, wherein variants in the variant set contain fewer than 5 substitutions.

126. (Previously presented) The method of claim 117, wherein the new variant set comprises variants of the biopolymer that have one or more substitutions at one or more positions that are not encompassed by the biopolymer sequence space of step a).

127 - 128. (Cancelled)

129. (Previously presented) The method of claim 117, wherein variants in the new variant set differ by fewer than 5 substitutions from at least one biopolymer for which the property has already been measured.

130 – 132. (Cancelled)

133. (Previously presented) The method of claim 118, the method further comprising repeating steps b) through f), until a variant in the new variant set exhibits a value for the property that exceeds a predetermined value.

134. (Previously presented) The method of claim 133, wherein the predetermined value is a value that is greater than the value for the property that is exhibited by the biopolymer of interest.

135. (Previously presented) The method of claim 118, the method further comprising repeating steps b) though f), until a variant in the variant set exhibits a value for the property that is less than a predetermined value.

136. (Previously presented) The method of claim 135, wherein the predetermined value is a value that is less than the value for the property that is exhibited by the biopolymer of interest.

137. (Cancelled)

138. (Withdrawn) The method of claim 1, wherein the modeling comprises least square regression, linear regression, non-linear regression, logistic regression, or partial least squares projection of latent.

139. (Cancelled)

140. (Previously presented) The method of claim 1, wherein the modeling step d) comprises:

computation of a neural network, computation of a Bayesian model, a generalized additive model, a support vector machine, machine learning, or classification using a regression tree using, as input to the modeling, (i) the one or more substitutions at the one or more positions of the biopolymer of interest represented by the variant set and (ii) the property measured for the variants in the variant set, and

obtaining, as output to the modeling, a predicted value for the property.

141. (Withdrawn) The method of claim 1, wherein the modeling step d) comprises boosting or adaptive boosting.

142 – 146. (Cancelled)

147. (Previously presented) The method of claim 117, wherein the plurality of positions and the one or more substitutions for each respective position in the plurality of positions are identified using a plurality of rules; and wherein

the contribution of each respective rule in the plurality of rules to the biopolymer sequence space is independently weighted by a rule weight in a plurality of rule weights corresponding to the respective rule; and

the method further comprises, prior to the defining of a new variant set step e), the steps of:

adjusting one or more rule weights in the plurality of rule weights based on a comparison, for each respective substitution at each position in the plurality of positions in the variant set, (i) a value derived for the respective substitution at each position in the plurality of positions from the sequence-activity relationship, and (ii) a score assigned by the plurality of rules to the respective substitution at each position in the plurality of positions; and

repeating the identifying step using the rule weights, thereby redefining the plurality of positions and, for each respective position in the plurality of positions, redefining the one or more substitutions for the respective position; and wherein

the defining of a new variant set step e) further comprises redefining the variant set to comprise one or more variants each having a substitution in a position in the redefined plurality of positions not present in any variant in the variant set selected by the initial selecting step b).

148. (Previously presented) The method of claim 117 wherein

the modeling a sequence-activity relationship d) further comprises modeling a plurality of sequence-activity relationships, wherein each respective sequence-activity

relationship in the plurality of sequence-activity relationships describes the relationship between (i) one or more substitutions at one or more positions of the biopolymer of interest represented by the variant set and (ii) the property measured for all or the portion of the variants in the variant set; and

the defining the variant set e) comprises redefining the variant set to comprise variants that include substitutions in the plurality of positions that are selected based on a combination function of the plurality of sequence-activity relationships.

149. (Cancelled)

150. (Previously presented) The method of claim 1, wherein the biopolymer of interest is a polypeptide, a polynucleotide, a small inhibitory RNA molecule (siRNA), or a polyketide.

151. (Withdrawn) The method of claim 1, wherein the biopolymer of interest is a protein kinase, a protein phosphatase, a protease, a receptor, a G-protein coupled receptor, a cytokine, a growth factor or an antigen from an infectious pathogen.

152. (Previously presented) The method of claim 1, wherein the biopolymer of interest is a cytochrome P450, a lipase, an esterase, a peptidase, a transferase, a polymerase, or a depolymerase.

153. (Previously presented) The method of claim 1, wherein the plurality of positions comprises five or more positions.

154. (Previously presented) The method of claim 1, wherein the plurality of positions comprises ten or more positions.

155. (Previously presented) The method of claim 119, wherein the plurality of rules comprises five or more rules.

156. (Previously presented) The method of claim 119, wherein

(A) the identifying combines a score from each rule in a plurality of rules thereby forming a cumulative score for each respective substitution at each position in the plurality of positions by summing the score from each rule in the plurality of rules for each respective substitution at each position in the plurality of positions, and

(B) the cumulative score for each respective substitution at each position in the plurality of positions is rank ordered.

157. (Previously presented) The method of claim 156, wherein the combining comprises adding (i) a first score from a first rule in the plurality rules and (ii) a second score from a second rule in the plurality rules for the variant of a biopolymer of interest.

158. (Previously presented) The method of claim 156, wherein

(A) the identifying combines a score from each rule in the plurality of rules thereby forming a cumulative score for each respective substitution at each position in the plurality of positions wherein the forming comprises multiplying (i) a first score from a first rule in the plurality rules and (ii) a second score from a second rule in the plurality rules for each respective substitution at each position in the plurality of, and

(B) the cumulative score for each respective substitution at each position in the plurality of positions is rank ordered.

159. (Currently amended) The method of claim 1, wherein the selecting the first plurality of variants step b) comprises applying a monte carlo algorithm, a genetic algorithm, or a combination thereof, to construct the variant set, with the provisos that:

(i) each variant in all or portion of the variant set has a number of substitutions that is between a first value and a second value; and

(ii) a number of different pairs of substitutions collectively represented by the variant set is above a predetermined number.

160. (Previously presented) The method of claim 159, wherein the first value is two substitutions and the second value is twenty substitutions.

161. (Previously presented) The method of claim 159, wherein the first value is four substitutions and the second value is ten substitutions.

162. (Previously presented) The method of claim 159, wherein the predetermined number is one hundred.

163. (Previously presented) The method of claim 1 wherein

the measuring step c) comprises synthesizing all or the portion of the variants in the variant set, and wherein

the property of a variant in the variant set is an antigenicity of the variant, an immunogenicity of the variant, an immunomodulatory activity of the variant, a catalysis of a chemical reaction by the variant, a thermostability of the variant, a level of expression of the variant in a host cell, a susceptibility of the variant to a post-translational modification, a killing of pathogenic organisms or viruses resulting from activity of the variant or a modulation of a signaling pathway by the variant.

164 – 169. (Cancelled)

170. (Previously presented) The method of claim 159, wherein the predetermined number is thirty.

171. (Previously presented) The method of claim 1, wherein each variant in the first plurality of variants is selected on a predetermined basis.

172. (Previously presented) The method of claim 1, wherein the value quantifying the confidence with which a substitution in the one or more substitutions of a position in the one or more positions of the biopolymer of interest contributes to the measured property is determined by the method of:

(i) calculating a plurality of sequence activity relationships, wherein each sequence activity relationship in the plurality of sequence activity relationships is calculated using the measured property of an independent subset of the variant set;

(ii) calculating, for each sequence activity relationship in said plurality of sequence activity relationships, a value for the contribution to the measured property by the substitution in the position; and

(iii) calculating a confidence for the value for the contribution to the measured property by the substitution in the position using each said value computed in said calculating step (ii).

173. (Previously presented) The method of claim 1 implemented on a computer.

174. (Previously presented) A computer program product encoding instructions for implementing the method according to claim 1.

175. (Previously presented) The method of claim 1 wherein the function f is a linear combination of the x_i and the sequence-activity relationship has the form:

$$Y= w_1x_1+ w_2x_2,+\ldots + w_ix_i.$$

176. (Withdrawn, Currently amended) The method of claim [[17]] 175 wherein a respective x_i in the sequence-activity relationship is a descriptor of a substitution or a combination of substitutions and wherein the substitution or combination of substitutions is selected for the new variant set for the biopolymer of interest when the weight w_i corresponding to the respective x_i is positive.

177. (Withdrawn) The method of claim 176 wherein the weight w_i corresponding to the respective x_i is at least one standard deviation above neutrality.

178. (Withdrawn) The method of claim 176 wherein the substitution or combination of substitutions has been tested at least three times.